

Improved Disparity Map Analysis Through the Fusion of Monocular Image Segmentations

Frederic P. Perlant and David M. McKeown

Digital Mapping Laboratory
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213¹

Abstract

It is generally believed that the detailed analysis of remotely sensed imagery requires the extraction of a variety of partial image domain cues coupled with the use of *a priori* or contextual information. In some cases there are fundamental limits to the variety and type of information that may be extracted from a single image or stereo pair. However, in most cases a sufficient variety of cues can be extracted; the major issue is in how to utilize disparate scene cues to achieve a more complete and accurate overall scene interpretation.

The focus of this paper is to examine how estimates of three-dimensional scene structure, as encoded in a scene disparity map, can be improved by the analysis of the original monocular imagery. This paper describes the utilization of surface illumination information provided by the segmentation of the monocular image into fine surface patches of nearly homogeneous intensity to remove mismatches generated during stereo matching. These patches are used to guide a statistical analysis of the disparity map based on the assumption that such patches correspond closely with physical surfaces in the scene. Such a technique is quite independent of whether the initial disparity map was generated by automated area-based or feature-based stereo matching.

We present stereo analysis results on a complex urban scene containing various man-made and natural features. This scene contains a variety of problems including low building height with respect to the stereo baseline, buildings and roads in complex terrain, and highly textured buildings and terrain. We demonstrate the improvements due to monocular fusion with a set of different region-based image segmentations. Finally, we discuss the generality of this approach to stereo analysis and its utility in the development of general three-dimensional scene interpretation systems.

¹This research was primarily supported by the U.S. Army Engineer Topographic Laboratories under Contract DACA72-87-C-0001 and partially supported by the Air Force Office of Scientific Research, under Grant AFOSR-89-0199, and by the Defense Advanced Research Projects Agency, DoD, through DARPA order 4976, and monitored by the Air Force Avionics Laboratory Under Contract F33615-87-C-1499. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the U.S. Army Engineer Topographic Laboratories, the Air Force Office of Scientific Research, the Defense Advanced Research Projects Agency, or of the United States Government.

1. Introduction

One common problem for systems that interpret multiple sources of sensed data is the fusion of partial results from a variety of sources. This problem appears under many guises. For example, given a set of different scene descriptions generated from a single image using a variety of image analysis techniques, how does one intelligently combine such partial information? [8]. The introduction of additional sensor types, temporal imagery, and multiple-look imagery create dimensions along which information fusion must be performed; as such, the complexity of the problem can increase. In some cases, increased amounts of data provide improved information. This may not necessarily follow, however; complex systems having different sources of error may not reinforce correct partial interpretations nor refute incorrect ones.

Thus, the key issue is the integration of many different sources of partial information. In computer vision (and in particular, three-dimensional scene analysis), the goal is to generate an *interpretation* of the scene that is as close as possible to the actual scene imaged. Such an interpretation can include the delineations and heights of buildings, a digital elevation model, and the centerline and width of roads in a transportation network. Our belief is that no individual computer vision technique can reliably provide a complete scene reconstruction. To achieve good performance, we need to gather a variety of information, extracted by various processes from the imagery, and synthesize this disparate information into a consistent model. Figure 1-1 shows a possible structure for such a scene interpretation system.

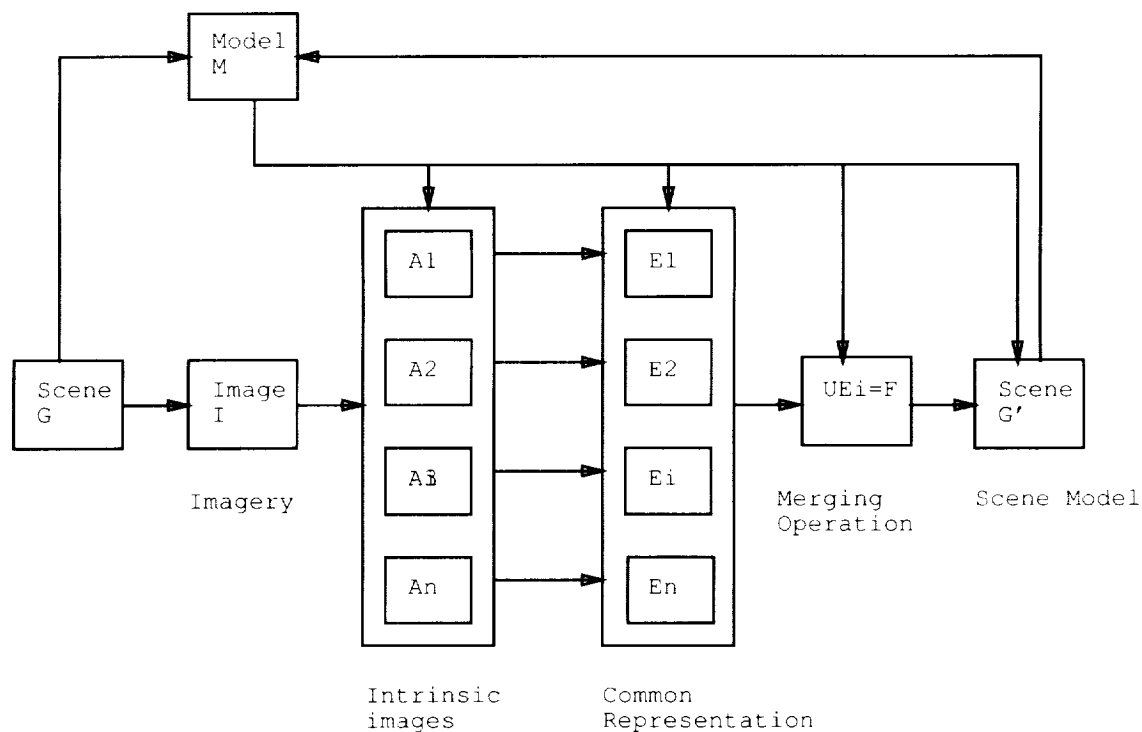


Figure 1-1: Data fusion in image analysis

From the three-dimensional scene (G) we generally acquire two-dimensional imagery generated by a variety of different sensors. For example, a stereo pair of intensity images would represent such an imagery. As is well understood, the problem of interpreting the two-dimensional image (I) as a three-dimensional scene is underconstrained. In certain cases, we

may have access to high-level knowledge about the contents of the scene, or particular objects that can be found in the scene. Such knowledge can loosely be called a Model (M). For example, in the case of aerial imagery we may have knowledge about the sensor resolution, the general characteristics of the scene (airport, urban area, rural area), etc. From the representation (I), we try to extract features that will allow us to interpret the scene $\{A_i\}$. These features are typically segmentations, edge maps, disparity maps, intensity maps, and the like. These can be thought of as a set of intrinsic images and primitives for intermediate and high-level vision [1, 7]. In order to fuse the information embodied in these different "images", we need a common framework of representations (formed by the $\{E_i\}$). This framework needs to allow many, if not all, of the $\{A_i\}$ features to be represented. The utilization of a common representation makes information fusion simpler and allows the generation of an interpretation (F), which then allows the generation of our scene model (G'). This model can be used to iterate through the fusion process again in conjunction with extra knowledge about the scene obtained from (M). This initial interpretation of the scene can help in the extraction of features $\{A_i\}$, the transformation of the features in the common representation, the merging process, and even the generation of the scene model.

Depending on the interpretation of the scene for which we are looking, we may need a varying amount of information; in most cases, more information is generally desirable. For instance, many techniques extract most of the necessary information for scene interpretation from a single intensity image; such techniques are said to apply *monocular analysis*. It is possible to take advantage of stereo disparity, however, to obtain more information that may be useful for disambiguation of monocular interpretations. Techniques utilizing stereo imagery are said to apply *binocular analysis* or *stereo analysis*. Other information such as global constraints or world models can be useful for further interpretation and disambiguation, but we believe that stereo analysis is a necessary step towards a coherent interpretation of the scene.

In this paper we describe a technique to merge information extracted from aerial imagery using a common region-based representation and show how disparate scene cues can be integrated to achieve a more complete and accurate overall scene interpretation. In Section 2 we describe techniques to improve the accuracy of a stereo disparity map using a single segmentation of the left intensity image of a stereo pair. Thus, we are able to recover from mismatches generated during stereo matching by re-utilizing the intensity image that was originally used in the matching process. In Section 3 we discuss some experimental results on disparity refinement and describe techniques that allow for the integration of additional scene segmentations to provide for a more robust refinement process. Finally, in Section 4 we give some future directions of this work in building extraction and built-up area analysis and speculate on how these techniques could be integrated into a more general three-dimensional scene interpretation system.

2. One approach to information fusion

In our research we utilize scene domain cues derived from monocular analysis and stereo analysis of left/right stereo image pairs. In the case of monocular analysis, one source of information is a region based segmentation of the left or right image. In the case of stereo analysis, our cues are primarily disparity maps derived from area-based and feature-based stereo matching algorithms. These image-based cues are different manifestations of man-made structures and terrain surfaces in the scene. In the case of three-dimensional reconstruction, we can make the assumption that the scene is composed of surfaces whose information content is primarily in terms of surface orientation and radiometry. Under these assumptions, we will see how estimates of three-dimensional scene structure (as encoded in a scene disparity map) can be

improved by the analysis of the original monocular imagery.

We have two sources of information that can be viewed as different representations of the physical surfaces found in the scene: disparity maps resulting from different stereo matchers providing the heights of the surfaces in the scene and the initial intensity images representing the radiometric properties of the surfaces in the scene. Figures 2-1 and 2-2 show an example of "initial" data used for these data fusion experiments. Figure 2-1 is a high resolution aerial image containing a variety of buildings with complex shapes, typical of an industrial area. Figure 2-2 is a disparity map derived using a feature-based stereo matching algorithm. These images are two of the many possible intrinsic images, $\{A_i\}$, in our general framework. It is important to note that, as in the intrinsic image paradigm, these two sources of information are "registered". That is, there is a pixel-by-pixel correspondence between points in the intensity image and points in the disparity map. In some many cases one issue complicating the use of multi-source information is the accurate registration or correspondence between the information sources themselves.



Figure 2-1: DC38008 industrial left intensity image

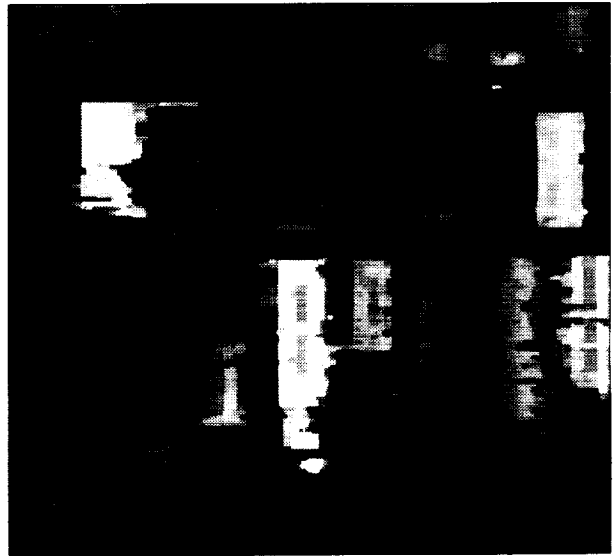


Figure 2-2: S2 left disparity map

An intensity image, subject to sampling and digitization errors, poses difficulties for monocular analysis techniques such as segmentation. On the other hand, most stereo matching algorithms are fooled by different variations in the stereo pairs, which cause mismatches in the disparity maps. The mismatches in disparity maps primarily result from geometric and radiometric differences in the left and right images, rather than local digitization or sampling errors in the intensity images. Thus, it is possible to use information from the intensity images to reduce the number of mismatches introduced by stereo matching processes.

2.1. Region based interpretation

Our approach utilizes surface illumination information, provided by the segmentation of the monocular images into fine surface patches of nearly homogeneous intensity, to remove mismatches generated during stereo matching. First, we segment the intensity image into uniform intensity regions. These regions correspond to approximately planar surfaces in the image. We assume that the orientation and surface material are the primary factors for the

radiometry of the image. Under these assumptions, uniform image radiometry is produced by a planar surface, of a certain orientation and material, in the scene.

These surfaces should have continuous linear disparity values (i.e., the disparity values of these regions are represented by continuous linear functions). Since the disparity map contains some noise, however, most of the regions segmented in the intensity image have disparity functions that are neither linear nor continuous. Ideally, we would like to approximate the actual disparity functions over the uniform intensity regions by the appropriate linear functions.

The problem of approximating a surface in three-dimensional space to a reasonable planar surface is a difficult one; we approximate such surfaces by horizontal surfaces. Then, the disparity values for each region will be the same for each pixel, and the problem is reduced to the selection of the best value for the heights of these surfaces. The general problem is that of locating of the surface which satisfies the equation

$$ax+by+cz+d=0$$

Given (x,y) , we should be able to obtain

$$z = (-ax-by-d)/c$$

We assume here that $z' = -d'/c'$ only. Then the problem is to find $(-d'/c')$ that best fits the surface so that

$$ax+by+c*(-d'/c')+d \sim 0$$

or to find z' so that $z-z'$ would have a minimal value over the region (this can be the weighted mean of the z distribution or the most 'representative' value of the z distribution). In other words, we need only select a single disparity value for each region. Since we are using an over-segmentation of the image, a piecewise planar disparity map gives a good approximation of the relief in the scene. Furthermore, since we are interested in building extraction in aerial images, this approximation will be adequate.

This region-based interpretation has been developed for two different applications. We show how this approach can support information fusion from different segmentations and well as across multiple disparity estimates based upon a local decision making evaluation. In Section 3.1 we describe how improved disparity maps may be obtained by correcting the mismatches produced by stereo matchers and by refining the disparity discontinuities. In Section 3.2 we will extract buildings from the scene using the height information in these disparity maps.

2.2. Intensity Segmentation Techniques

The general scene segmentation problem is, of course, a very difficult one and has a long history in image processing and computer vision. There are no universal segmentation techniques that work well across a variety of imagery and tasks. Such low level algorithms typically differ in their approaches; they may utilize intensity-based, area-based, or edge-based techniques. Some systems combine these techniques into hybrid algorithms. We have concentrated on those segmentation methods that produce (nearly) uniform intensity regions because we wish to detect those image regions that correspond to oriented surface patches in the scene. We utilize a region segmentation algorithm based upon the histogram splitting paradigm [6] and a region growing algorithm [9] which takes into account edge strength and shape criteria [4]. Interestingly, while neither of these methods give completely satisfactory segmentation results, they provide good over-segmentations that rarely merge object/background boundaries. Both techniques will also provide different segmentations based upon modification

of a small set of parameters. In our experiments we generated three scene segmentations; two by using different parameters for histogram selection, and one by using region growing. These segmentations provided the basis for our work in intensity/disparity fusion, the goal of which was to produce an improved three-dimensional scene interpretation.

Figures 2-4 - 2-6 show examples of these segmentations on the DC38008 industrial left intensity image. We ran the experiments on smoothed images (Figure 2-3) to remove intensity noise.

2.2.1. Machineseg

One of the major difficulties with region growing techniques in complex scenes is the difficulty in determining automatic stopping conditions for the merging procedure. MACHINESEG [4] is a region growing system that tries to preserve edges between regions and stops the growing procedure when certain shape or spectral criteria are not satisfied inside the region. It adds a decision procedure to evaluate the effect of the next merge operation and either allows the merge to proceed or to be rejected. In the case of disparity map refinement, we want the regions to be sufficiently uniform that they could be treated as planar (or at least "soft") surfaces. We also limited the size of the generated regions so that very small regions could not be generated, as these could be considered noise or non-representative regions. As can be seen in Figure 2-4, since we are not considering the small region, our segmentation is not a complete partition of the image; it does, however, obtain most of the representative surfaces in the image.



Figure 2-3: Nagao filtered left image for DC38008



Figure 2-4: MACHINESEG segmentation on DC38008

2.2.2. Colorseg

This histogram splitting technique is based on the extraction of regions with limited intensity ranges (in other words, region of approximately uniform intensities). The technique searches for the peaks in the histogram of the image and segments the regions whose intensity values fall in windows around these peaks. The regions are then removed from the image and the process continues until all the pixels in the image have been removed. This process results in a segmentation composed of connected regions, each having an intensity range less than a certain threshold. This technique does not guarantee preservation of the edges (in particular, small edges) but it may ignore local noise with strong edges that other techniques will classify as

regions. As in the previous technique, we removed very small regions (less than 20 pixels) that could be considered as noise, for further processing.

In our experiments, we generated different segmentations with different segmentation techniques. For instance, using the colorseg technique we generated two segmentations of the images, one with "uniformity" defined as a maximum of 10 intensity levels inside the region (to tolerate sensor noise and allow for imperfect planar surfaces) and another with "uniformity" defined as a maximum of 20 intensity levels (to tolerate more noise). An estimation of the noise or the average intensity range for the surfaces in the image is a delicate problem, and the use of different segmentations to estimate the intensity range inside the regions does not necessarily increase the reliability of the process. It is thus important that we obtain different segmentations of the scene that are *not consistent*, such as those in Figures 2-5 and 2-6. The fusion of these data may overcome some of the inherent problems of a single segmentation since they provide different local evaluation contexts for disparity estimates in the scene. In the following sections we show how we can merge information using different intensity segmentations.

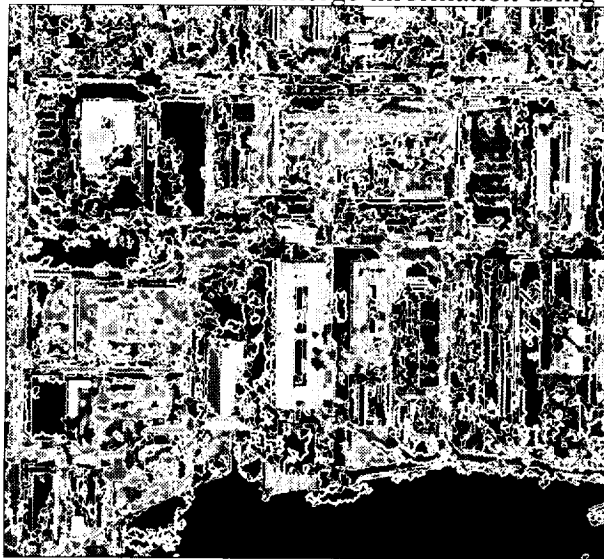


Figure 2-5: COLORSEG segmentation with 10 intensity levels sensitivity for DC38008

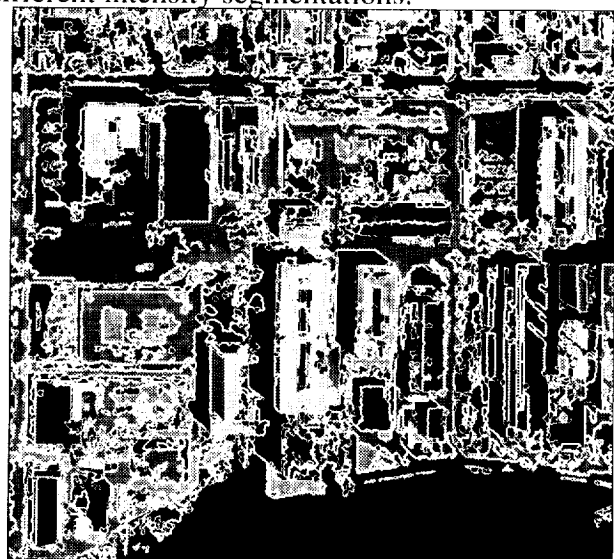


Figure 2-6: COLORSEG segmentation with 20 intensity levels sensitivity for DC38008

2.3. Disparity map results

Our initial height information for the industrial scene was derived using two different stereo matching algorithms. Given these sets of height information, which may or may not be reliable or unique, it becomes necessary to use a data fusion process in order to maximize the amount of useful information gained from these sets of height estimates.

We used 2 different matching techniques, one area_based (S1) and the other feature_based (S2). S1 uses the method of differences technique on neighborhoods of the image in hierarchical fashion [3, 5]. S2 performs a hierarchical matching of epipolar intensity scanlines in the left and right image [2]. The results of these stereo matching algorithms are different; S1 gives us a dense disparity map (i.e., a map containing a disparity value for each pixel in the image), while S2 gives us a sparse disparity map (i.e., a map containing a disparity value for those pixels corresponding to peaks or valleys in the intensity images).

Since we used uniform segmented regions that we assumed to be horizontal planes, a logical interpolation method for the sparse S2 disparity map is step interpolation. This produces a dense disparity map consisting of regions with uniform disparity values, which may be more easily integrated with a dense map produced by S1. Our fusion mechanism will have to correct mismatches in the S1 or S2 disparity maps and then choose the better unique disparity value for each pixel in the scene. It will have to merge very different disparity information, such as that shown in Figures 3-2 and 3-1, the two left disparity maps for the DC38008 scene.

3. Fusion Experiments

After different intensity segmentations and different disparity results were obtained, we applied a very simple fusion technique and developed a few experiments for the two applications under consideration. Most of the experiments have been performed for the disparity refinement process, but the results have been used for the building extraction process as well.

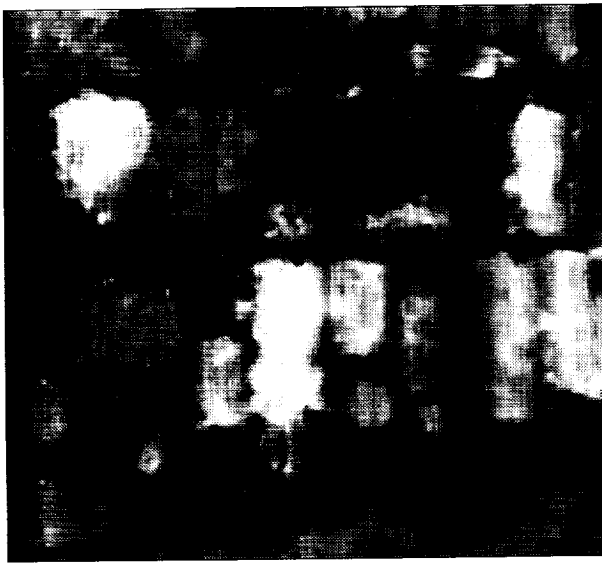


Figure 3-1: S1 left disparity
result for DC38008

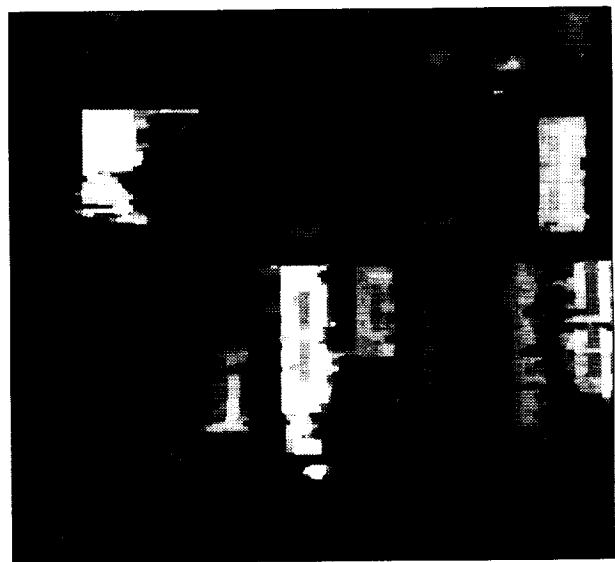


Figure 3-2: S2 left disparity
result for DC38008

3.1. Disparity refinement

In order to refine the disparity maps (i.e., to remove mismatches, improve disparity discontinuities and obtain the best height estimate for each point in the scene), several approaches have been explored:

- Disparity refinement using one segmentation
- Disparity refinement using several segmentations
- Disparity refinement using one segmentation and several disparity maps
- Disparity refinement using several segmentations and several disparity maps

3.1.1. Simple disparity refinement

In this first approach, a histogram is constructed for each segmentation region. The values of each histogram are the disparity values in each region. The most representative value of each histogram is then selected. In our case, this value was simply that of the highest peak in the histogram. We chose this value for two reasons. The step-interpolated S2 disparity maps result

in disparity histograms having only a few values, which correspond to real height values or matching noise. If the matching is reasonably robust, the noise will introduce local maxima in the histogram that will be smaller in magnitude than the best height estimate. Further, a typical region histogram for an S2 disparity map exhibits one or two large peaks and a few noise peaks that influence the average value of the histogram, making it less reliable as a representative value.

For non-horizontal regions and S1 results, the average disparity may suffice for a reasonable measure of the height of the region. A confidence score can be generated for these disparity values based on the characteristics of the histograms (and, conceivably, on the type of disparity map used as well as the nature of the region histograms). Finally, this disparity value is assigned to the entire region, under the assumption that it will be a better estimate of the height for the whole region. In most cases, this removes a large number of the mismatches, but whenever our initial assumptions about scene radiometry are not valid, our height estimates may differ from the correct height value.

We implemented this approach for each segmentation and disparity map and generated new disparity maps that were based on the initial intensity regions and disparity values. The pixels that were not considered during the segmentation were removed from these new disparity maps. Figures 3-3 and 3-4 show the results of the disparity improvement process for the different segmentations using the S2 disparity map, and Figures 3-5 and 3-6 show the results of the disparity improvement process for the S1 disparity map.



Figure 3-3: S2 left disparity
result for DC38008
improved using SEG10

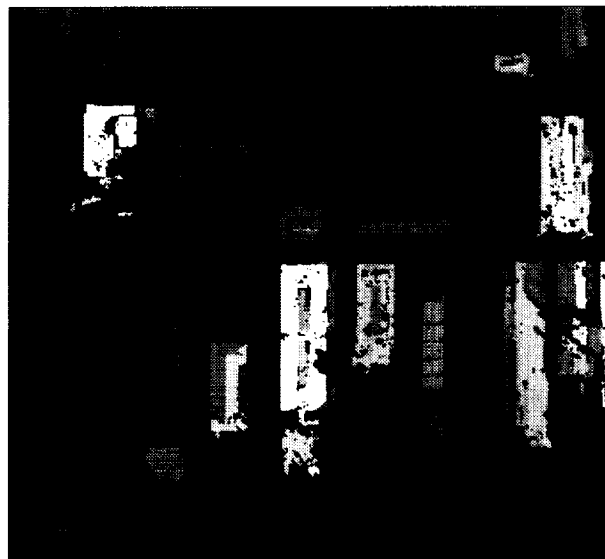


Figure 3-4: S2 left disparity
result for DC38008
improved using SEG20

It is worth noting that a common methodology is utilized among all of the approaches described in this section. A set of attributes is computed for each region in each segmentation. Among these attributes are the statistics for the disparity values inside a region, the best disparity value, and a confidence score for this value. This allows the computation to proceed at a symbolic level on a region-by-region basis.



Figure 3-5: S1 left disparity
result for DC38008
improved using SEG10



Figure 3-6: S1 left disparity
result for DC38008
improved using SEG20



Figure 3-7: S1 left disparity
result for DC38008
improved using the merging
of SEG10 and SEG20



Figure 3-8: S2 left disparity
result for DC38008
improved using the merging
of SEG10 and SEG20

3.1.2. Multi-segmentation disparity refinement

In the second approach, we can merge different height estimates, given different intensity segmentations (SEG10, SEG20) and then merging the results across the different segmentations. We refine the disparity estimate for each pixel by locating the intensity region to which it belongs, for each of the image segmentations. This list of regions can then be searched to obtain the disparity estimate attribute (computed for a given disparity map) as well as a confidence

score for this estimate. The confidence score is then used to select the best disparity value, which is then assigned to the pixel. Currently a simple decision is made to select the disparity value having the highest confidence score.

An attempt is made to maximize the score for each pixel in the entire image. This is done by selecting a disparity value in all of the regions resulting from the union of the segmentations. In other words, the segmentations were merged and the best height value was selected for each of these regions, by utilizing the confidence scores computed for each region. The scoring method currently in use takes into account information about the nature of the segmentation used.

In particular, higher confidences can be assigned to sufficiently large regions in a constrained segmentation such as SEG10 than to the equivalent regions in SEG20. Information of this nature must be incorporated in the confidence function for each segmentation region.

Figures 3-8 and 3-7 show the results of merging the SEG10 and the SEG20 segmentations for the S2 and the S1 disparity maps, respectively. Depending on the confidence scores of the disparity values selected for each segmentation, we were able to obtain improved disparity estimates for some of the regions. Comparing these results to Figures 3-3 and 3-4, disparity maps obtained with the simple method, we observe some of the failings of both approaches. The initial segmentations, in some cases, are under-segmented instead of over-segmented, resulting in the grouping of regions that should have been assigned different height estimates. Another factor is the confidence evaluation function for the regions of the segmentation, which only takes simple properties of the disparity histograms of each region into account.

3.1.3. Multi-Disparity Disparity Refinement

In this approach, several different disparity maps are merged using a single segmentation, looking for consistent areas across disparity maps. This approach is similar to the simple disparity improvement approach, except that we now attempt to select the best disparity value based on a set of differing confidence scores. The score established for each disparity map at each pixel should be dependent on the stereo matching algorithm used to generate the map, and should also take into account the nature of the possible mismatches resulting from each stereo matching technique.

The major problem with all of the refinement approaches discussed in this paper is the development of a reasonable confidence evaluation function for each set of data. Currently, confidence is evaluated by a scoring function that utilizes the standard deviation and the disparity range of the histogram for each region, as well as the size of the region. Ideally, this scoring function would also take into account the nature of the disparity map. As an initial experiment, we defined a similar scoring function for each disparity map and checked for disparity consistency across segmentation regions. In Figure 3-9, the areas where disparity values differ between S1 and S2 are marked in black, as we do not use any score difference information to select the most probable height value at this stage.

3.1.4. General Disparity Refinement

For the general case we can merge the results of different disparity maps and different segmentations and look for consistency across the results. The approach is similar to the multi-segmentation method; however, we should be able to add additional height hypotheses according to the different segmentations.

Again, the processes can be decomposed into two stages. The first stage will gather the

information and convert it into a common representation (i.e., region attributes). As an example, for each segmentation we should obtain a list of height estimates with scores associated with each of the different disparity maps we can use (S1 and S2). The second stage will attempt to merge this information by selecting the "correct" value from the available information, by comparing scores based on the nature and quality of the different pieces of information. If we can precisely evaluate the quality or confidence in the information, we should be able to maximize the amount of accurate data we merge from our different information sources.

There are still many experiments that have yet to be performed. In particular, experimentation needs to be done on merging the two different disparity values for the three different segmentations.

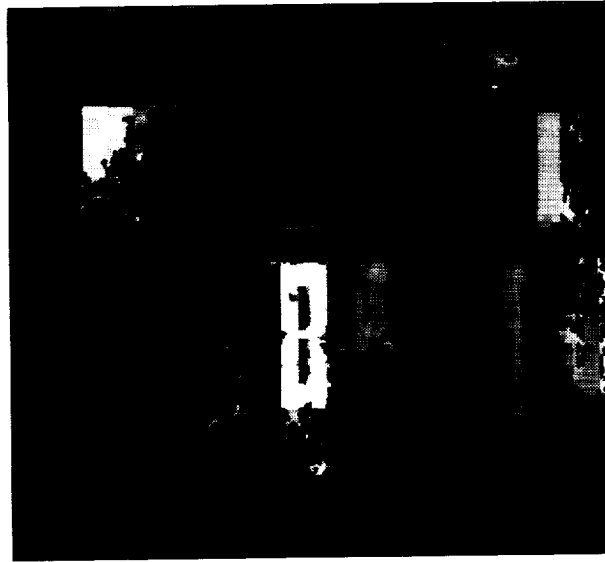


Figure 3-9: S1 left disparity
and S2 left disparity
merged using YAK

3.2. Building extraction

This second application of information fusion is an attempt to validate this region-based approach for scene interpretation. Using the previously described methods, we can obtain an estimate of the height of each of the composite regions in each segmentation. According to our representation of the scene, buildings are composed of a single intensity region or a group of intensity regions, and, in general, are higher than their surroundings. Therefore, regions representing parts of a building should be higher than their neighboring regions.

For each region, a list of its neighboring regions is constructed, and the disparity values for each of these regions are obtained. Then, a weighted histogram is computed that takes into account shared boundary length and disparity information. This weighted score is then compared with the height of the region to label the region as building structure or background terrain. This building extraction process can use either the initial disparity map or the refined disparity map.

A refinement process is used to group neighboring regions with the same height in order to obtain an intermediate segmentation containing fewer (and larger) consistent regions. This

grouping procedure merges connected regions having the same height to form a single region. This allows the building extraction process to use larger, and hopefully more consistent, disparity regions as a basis for the neighborhood disparity analysis. The quality of this analysis is again dependent on the accuracy of the disparity estimate, as in the previous fusion process.

Figure 3-10 shows the result of such an analysis. The white regions correspond to sections of buildings. The building extraction, as done by hand, is in Figure 3-11.



Figure 3-10: Building regions for DC38008 extracted using the merging of SEG10 and SEG20



Figure 3-11: Building regions for DC38008 extracted manually

The problem can be described as the use of "early" or "initial" information for which we do not have any confidence measures to construct a model. To perform this task, we must gather confidence about this information as computation proceeds in order to construct a three-dimensional interpretation of the scene. The building extraction process described here illustrates one facet of scene interpretation that can be performed within this framework.

4. Conclusions

We have described a set of fusion processes that allow us to improve the quality of disparity maps, and we have demonstrated the use of information fusion to improve disparity map analysis. We described a building extraction approach that utilized the fusion technique. The major feature of the information fusion technique described here is the definition of a common frame for information fusion. The representation framework (an intensity segmentation) can be used in conjunction with different types of intrinsic images. The approach developed here treats homogeneous intensity regions as surfaces, which allows three-dimensional information to be extracted readily.

Many research issues remain to be explored. The new disparity maps generated by the information fusion process contain regions which each have only one disparity value. In many cases, these unique values are not the best possible disparity estimates for the regions, and a refinement process may need to be invoked to correct these estimates. One approach might be to

use the new disparity map itself as input to a verification process which could refine disparity estimates for each pixel or for those regions with low confidence scores.

Other sources of information could be utilized at the refinement stage to further enhance the disparity map. One promising approach would be the use of left/right consistency, such as left/right matching of low confidence regions or local correlation for these regions. Again, it would be important to use as much information as possible, while conservatively adjusting or refining data based on its confidence scores. In the ideal situation, no additional information would refine the disparity estimates; it would merely verify the truth of the disparity map.

Many improvements can be obtained by the use of better segmentations and scoring functions, and by addressing the assumption that only flat horizontal surfaces are responsible for the imaged radiometry and by using a more sophisticated surface model such as non-horizontal planar surfaces or quadratic surfaces. Finally, it seems feasible that multispectral data could be integrated by similar techniques. The information fusion approaches described here provide a means for data integration that may prove useful in other aspects of scene interpretation.

5. Acknowledgments

We would like to thank our colleagues in the Digital Mapping Laboratory for an interesting and congenial working environment. Jeff Shufelt provided helpful comments and suggestions on an earlier draft of this paper.

References

- [1] Barrow, H. G. and Tenenbaum, J. M.
Recovering Intrinsic Scene Characteristics from Images.
In Hanson, A. R. and Riseman, E. M. (editor), *Computer Vision Systems*, pages 3-26.
Academic Press, New York, 1978.
- [2] Hsieh, Y., Perlant, F., and McKeown, D. M.
Recovering 3D Information from Complex Aerial Imagery.
In *Proceedings: 10th International Conference on Pattern Recognition, Atlantic City, New Jersey*, pages 136-146. June, 1990.
- [3] B. D. Lucas.
Generalized Image Matching By The Method of Differences.
PhD thesis, Carnegie Mellon University, July, 1984.
- [4] McKeown, D.M., Denlinger, J.L.
Map-Guided Feature Extraction from Aerial Imagery.
In *Proceedings of Second IEEE Computer Society Workshop on Computer Vision: Representation and Control*, pages 205-213. May, 1984.
- [5] McKeown, D.M., McVay, C.A., and Lucas, B. D.
Stereo Verification In Aerial Image Analysis.
Optical Engineering 25(3):333-346, March, 1986.
- [6] Ohlander, R., Price, K., and Reddy, D. R.
Picture Segmentation Using a Recursive Region Splitting Method.
Computer Graphics and Image Processing 8:313-333, 1978.

- [7] Poggio, T. et. al.
The MIT Vision Machine.
In *Proceedings of the DARPA Image Understanding Workshop*, pages 177,198. Morgan Kaufmann, April, 1988.
- [8] Shufelt, J., McKeown, D. M.
Fusion of Monocular Cues to Detect Man-Made Structures in Aerial Imagery.
In *Proceedings: IAPR Workshop on Multisource Data Integration in Remote Sensing*.
June, 1990.
- [9] Yoram Yakimovsky.
Boundary and Object Detection in Real World Images.
Journal of the ACM 23(4), 1976.

